

Modern Statistical Computing

Seminar #1

January 19, 2023

Homework part due on January 26, 2023 (3pm)

The idea of this seminar is to pull data from [Eurostat](#) on demographic variables, do some data wrangling, and ultimately generate some plots. We will fetch data at the geographic level of NUTS2 regions (e.g. Catalonia in Spain, to get a general sense of NUTS2 regions you can look at maps [here](#)).

Exercises

Building the Dataset

1. Go to [Eurostat](#) and search for data at the **NUTS2 regional level** on the following regional demographic statistics:
 - population
 - population density
 - changes in population (naturally and via migration)
 - fertility rates

Make sure you understand how these variables are constructed before moving on.

2. Pull data on these variables for the years **2011-2019**. As you will later work with data for Spain and a country of your interest, you can either pull data for all countries and filter later, or filter directly when downloading the data.
To pull the data you can proceed as follows:
 1. Use the eurostat package ([documentation](#)) and specifically its function `get_eurostat` (using the `tablenames` you have found in 1.) to pull the data directly into R
 2. Manually download the data from eurostat (e.g. as .csv files) and then read it into R.
3. Generate a NUTS2 \times year dataset with the demographic variables from above as columns. To map NUTS2 codes to region names, you can further merge your dataset with the `nuts2_overview.xlsx` helpfile provided on Aula Global.
4. Aggregate the NUTS2 \times year dataset to the nation \times year level with information on the population stock and fertility rate. For the fertility rate, use population weights when aggregating from the NUTS2 to the national level.

Describing Demographic Trends

From now onward, you will focus on (NUTS2 regions in) Spain and one country of your interest.¹

1. Compute the *national* population growth rates (in %) and plot them separately by country from over time. In a separate graph, plot the countries' fertility rates over time. Do the two countries follow similar trends in their population dynamics?

¹Some of the countries in the dataset have missing values in the first years of the selected time period (e.g. Albania until 2016). It might thus make sense to select a country with no missings throughout the entire sample period.

2. Merge the national annual growth rates to your NUTS2 \times year dataset. Create a variable that captures whether a region grows faster or slower than the national average for a given year. Which regions growth particularly more (or less) than their country over the years in your sample?
3. Next, we would like to understand whether migration or natural population growth contribute more to local population dynamics. Generate, separately by country, a graph that shows for each year the (mean) contributions of migration and natural growth to population growth. Are the patterns different when looking separately by regions that grow more/less than their country?

Migrant Concentration (Homework, due: 26-01-2023, 3pm)

Lastly, we would like to study whether migrants locate in different regions compared to natives. We will use data on population stocks by country of birth from the 2011 census rounds which you can find under Population and social conditions on [Eurostat](#). Again our regional units will be at the NUTS2 level. As before, focus on Spain and another country of your interest.

1. Pull data on population by group of country of birth and educational attainment (tablename: cens_11_cobe_r2). Build a NUTS2-level dataset with information on the total stock of natives (c_birth = NAT) and migrants (c_birth = FOR) across education groups, and by education groups. In terms of education groups, aggregate the ISCED categories used from Eurostat to four broader categories (no education, primary, secondary and tertiary education).
2. Generate a migrant concentration variable for each region r in country c as follows

$$\text{Migrant Concentration}_{rc} = \frac{\text{Migrants}_{rc} / \text{Migrants}_c}{\text{Natives}_{rc} / \text{Natives}_c}$$

Which are the regions for each country in which migrants concentrate the most?

3. Merge the population density from your NUTS2 \times year dataset from the second part to each region in 2011 (the year of the census). Assess visually, whether migrants concentrate more in more dense areas than natives. Does this hold for all education groups?